Validation of the Spanish version of the Childhood Executive Functioning Inventory (CHEXI) in 4-5 year-old children

Almudena Giménez[1*], Miguel López-Zamora[2], Oscar Vila[3], Auxiliadora Sánchez[1], and Lisa B. Thorell[4]

1 Universidad de Málaga. (Spain)
2 Universidad de Granada. (Spain)
3 UNED. (Spain)
4 Karolinska Institutet. (Sweden)

Título: Fiabilidad y validez de la versión española del "Childhood Executive Functioning Inventory" (CHEXI).
Resumen: *Contexto:* El Childhood Executive Functioning Inventory (CHEXI) se diseñó para detectar déficit de Funciones Ejecutivas (FE) en contextos de la vida diaria. El propósito del presente estudio era validar la versión española del CHEXI para niños de 4 y 5 años. *Método:* Se obtuvieron las valoraciones de los padres y las puntuaciones en test de laboratorio en una muestra de 445 niños de 4 años (196 niñas, 249niños) y otra de 459 niños de 5 años (208 niñas, 251 niños). Un año más tarde se volvieron a recoger las valoraciones de los padres en CHEXI. *Resultados:* Se replicó la estructura de dos factores de CHEXI (Memoria de Trabajo e Inhibición), obteniendo una alta consistencia interna y estabilidad temporal. Los niños de 4 años eran valorados con mayor déficit en FE que los niños de 5 años. Los niños alcanzaban puntuaciones más elevadas de déficit de FE. Sin embargo, no aparecieron diferencias significativas de género en las pruebas conductuales. Finalmente, se encontró una baja asociación entre las valoraciones de CHEXI y las puntuaciones de las pruebas conductuales. Este resultado sugiere que los inventarios y las medidas conductuales evalúan diferentes aspectos de las FE. *Conclusiones:* La versión española de CHEXI ofrece un instrumento válido para evaluar las FE en niños de 4 y 5 años.
Palabras clave: Inventario de FE. FE en niños de prescolar. Asociación entre las puntuaciones de FE en inventarios y pruebas conductuales.

Abstract: *Background:* The Childhood Executive Functioning Inventory (CHEXI) was designed to identify EF deficits in daily life contexts. The aim of the present study was to validate the Spanish version of the CHEXI in 4-5-year-old children. *Method:* Parents' ratings of the CHEXI and laboratory tests of EF were investigated in two large samples of 445 children aged 4 years (196 girls, 249 boys) and 459 children aged 5 years (208 girls, 251 boys). CHEXI ratings were collected again after one-year. *Results:* The two-factor structure of the CHEXI (i.e., Working Memory and Inhibition) was replicated, showing high internal consistency and temporal stability. The 4-year-olds were reported to have higher EF deficits than 5-year-olds have. Boys were rated as having higher EF deficits than girls were. However, gender differences were not significant contrasting performance on EF tasks. Finally, associations between CHEXI ratings and EF tests were weak, suggesting that EF tests and EF ratings capture different aspects of EF. *Conclusions:* The Spanish CHEXI provides a suitable instrument to assess EF in 4-5-year-old children.
Keywords: EF inventory. EF in preschool children. EF ratings and EF tests associations.

Introduction

Executive function (EF) is a construct defined as a set of tightly related high order cognitive skills involved in self-regulation needed to perform goal-directed behaviours (Miyake and Friedman, 2012). EF enables the mental manipulation of ideas, novel information management, inhibition, and concentration during the execution of complex tasks. According to Diamond (2013, 2020), EF comprises three core components: (a) inhibition – the ability to control attention and to refrain from responding to non-relevant stimuli, (b) working memory (WM) – the ability to hold and incorporate new information in memory, and (c) flexibility/shifting – the ability to go back and forth or to redirect attention according to task demands. More recently, García-Madruga, Gómez-Veiga, and Vila (2016) included a fourth core EF: focusing and sustaining attention. Past research has shown a strong association between early EF competence and later outcomes along adolescence and adulthood in many aspects of daily life (Best, Miller, & Jones, 2009). In addition, EF deficits have been observed in children with neurodevelop-

mental disorders (Willcutt et al., 2005). Due to the importance of early executive functioning in child development, it is necessary to have reliable and valid measures of EF in young children. The overall aim of the present study was therefore to investigate the psychometric properties of the Child Executive Functioning Inventory (CHEXI), a scale designed to assess EF in daily life contexts, for Spanish 4–5-year-old children.

Early identification of EF deficits

The assessment of EF is of special interest at the start of schooling. This age period is marked by dramatic increases in the efficiency of EF related to the maturation of frontal lobes (Romine & Reynolds, 2005). Increases in cognitive control are accompanied by better performance on EF measures. Thus, tasks that are extremely difficult at age 3 are easily solved by age 6 (Carlson, 2005; Diamond, Kirkham, & Amso, 2002). Furthermore, studies have pointed to continuity in the development of EF. For example, Friedman et al. (2007) found that participants who had better self-restraint when toddlers, achieved higher outcomes on inhibition, updating, and shifting tasks at 17 years. Another reason for early assessment is that strong EF competence in childhood has been linked positively to preparedness to school (Shaul & Schwartz, 2014), later academic achievement (e.g., Best, Mil-

ler & Naglieri, 2011; Blair & Diamond, 2008), and that children with good inhibitory control were less vulnerable to fail in social and learning contexts, and were more successful in life (Moffitt et al., 2011). Finally, this is the time to observe the first signs of difficulties that will affect later development (Isquith, Gioia, & Espy, 2004; Sjöwall, Bohlin, Rydell, & Thorell, 2017). In addition, due to the quick development of EF at this age, small (non-clinical) efficiency lags can also be observed in typically developing children (Thorell, & Catale, 2014). These results indicate the need of reliable instruments to assess preschoolers' EF functioning in daily life contexts.

Rating scales are appropriate for this purpose because they make it possible to evaluate the child's behaviour in natural contexts (i.e., they have high ecological validity). Inventories have two clear advantages over laboratory measures. First, they are easy to administer and to score, therefore cost-efficient. Second, the information gathered by inventories refers to a long period of time, while the measures obtained by the tasks are limited to a specific point in time.

The Childhood Executive Functioning Inventory (CHEXI)

The CHEXI (Thorell & Nyberg, 2008) was designed to focus specifically on deficits in WM and inhibition, without including items that are closely aligned with the symptom criteria for Attention Deficit Hyperactivity Disorder (ADHD; e.g., "is impulsive" and "has a short attention span"). The CHEXI consists of 24 items that describe daily life behaviours with a colloquial expression and concrete examples to make the items easy to understand for parents and teachers. The CHEXI can be completed in only 5 minutes what makes it suitable both as a screening tool for identifying EF deficits in children with special needs (for example those with neurodevelopmental disorders), as well as for rating EF among normally developing children in an educational setting or for research purposes.

The fact that the CHEXI can be downloaded free of charge (www.chexi.se) in different languages may have contributed to its quick dissemination since it first publication in Swedish. Validation studies have been conducted with French populations in France (Catale, Lejeune, Merbah, & Meulemans, 2013), and Belgium (Catale, Meulemans, & Thorell, 2015); with Portuguese speakers in Brazil (Tonietti, Martins, de Almeida, & Gotuzo, 2017), and with American English participants in the USA (Camerota, Willoughby, Kuhn, & Blair, 2018). Both the original validation study (Thorell & Nyberg, 2008) and other language versions of CHEXI have generally found a two-factor structure – WM and Inhibition – to be the best fitting model to the data. Only the Brazilian version showed a distribution of the items closer to a one factor structure that the authors attributed to that the EF components could be less differentiated in their 4-year-old participants (Tonietti et al., 2017). Previous research has also shown adequate reliability in 5-6 year old children (Thorell &

Nyberg, 2008), as well as for older children (Catale et al, 2015), and for the versions in different languages cited above. Finally, the CHEXI has been shown to be able to discriminate between children with ADHD and typically developing controls, with overall classification rates ranging between 84-94% (Catale et al., 2015; Thorell, Eninger, Brocki, & Bohlin, 2010). Taken together, these findings provide convincing evidence of the utility of the CHEXI as a screening tool to assess EF in children.

One way of evaluating the internal structure of the CHEXI ratings is to examine the extent to which they are related to performance-based measures. However, the majority of research has found weak associations. In a study including 6-year-old Swedish children, Thorell and Nyberg, (2008) reported moderate correlations between CHEXI parents' and teachers' ratings and the scores on a measure of inhibition (i.e., the go/no-go task) and on a word span task to assess WM. Furthermore, correlations were not restricted to the tasks designed to assess the skills corresponding to each specific factor. (i.e., both parents' and teachers' ratings on the memory factor were significantly associated with the inhibition task, and teachers' ratings on the inhibition factor correlated with word span). Other study with a sample of 5-7-year-old Belgian children (Catale et al., 2013) also failed to find relations between CHEXI ratings and different EF tasks and the review by Toplak, West, and Stanovich (2013) showed that weak relation between EF tests and EF ratings have been demonstrated also for other instruments. The authors emphasized that this should be taken as evidence that tests and ratings are measuring partly different constructs, rather than as a sign of poor validity and that both tests and ratings should be used as a complement to one another.

Aim of the present study

The CHEXI has been previously used in a cross-cultural study with 6 to 11-year-old Spanish children (Thorell, Veleiro, Siu, & Mohammadi, 2013). There is also an adult version, the Adult Executive Function Inventory (ADEXI; Holst & Thorell, 2018) already validated in Spain (García-Villamisar, Jodra-Chuan, Saez, & Thorell, 2020). However, except for split-half reliability and relations to academic achievement, the psychometric properties of the Spanish version of the CHEXI have not yet been examined. In addition, no previous study has examined the Spanish CHEXI in children below school age. This is a serious limitation as the CHEXI is frequently used in both research and within clinical settings in Spain. The overall aim of the present study was therefore to validate the Spanish version of the CHEXI in a preschool population. More specifically, the following issues were addressed:
1. To what extent the previously established two-factor structure of the CHEXI can be replicated for the Spanish CHEXI in preschool children.
2. Temporal stability in CHEXI parent ratings collected one year apart.

3. Effects of age and gender in executive functioning assessed using the Spanish CHEXI.
4. Associations between parent ratings using the Spanish CHEXI and EF laboratory tests.

We expected to find evidence of age differences as previous research has shown that EF abilities have a quick developmental progression during the preschool age (Diamond, 2006). In line with previous studies examining gender differences for CHEXI ratings (Camerota et al., 2018; Thorell et al., 2013), we expected boys to receive higher scores (i.e., poorer executive functioning) on the CHEXI compared to girls. With regard to associations between CHEXI parent ratings and EF tasks, we expected to find significant, although modest, correlations as previous studies indicated that both measures tap partially different constructs (Toplak et al., 2013).

Method

Participants and procedure

The present study used data from two different samples: a sample of 445 children (196 girls, 249 boys) in the 2nd year of preschool (mean age = 4.37 years SD = .291), and a sample of 459 children (208 girls, 251 boys) in the 1st year of primary school (M = 5.46 years SD = .284). Thus, the mean age difference between both samples was 1.09 years. The samples were recruited from 30 public schools in the province of Malaga (Spain). According to the classification norms of the region, 25.3% came from low, 35.9% from medium, and 38.9% from high SES families. Teachers were interviewed to make sure that none of the children had an intellectual disability or any severe deficits in sensory or motor functioning. Parents completed the CHEXI at home. A selection of parents (n = 115; 59 4-y-o; 56 5-y-o children) were asked to complete the CHEXI a second time one year later. The laboratory tests were completed individually in a room of the school the participants normally attended. Children were seated in front of a computer screen. Then, the examiner explained the tasks and instructed them in the use of the mouse. Tests were administered in a randomized order in a single session. The examiner paused when the child looked tired, what made the session vary in duration between 45-60 m.

Materials

Childhood Executive Functioning Inventory (CHEXI)

The CHEXI includes 24 items describing relatively common behaviours in different contexts that tap two underlying factors: WM and Inhibition (Thorell & Nyberg, 2008). Responders (parents or teachers) are requested to rate the child's behaviour on a five-point Likert scale ranging from 1 (definitely not true) to 5 (definitely true). Higher ratings are indicative of poor executive functioning.

Performance-based measures of executive functioning

For the present study, three EF tasks, originally designed for children from eight years, were adapted to be applied to pre-schoolers. The tasks and the adaptation are described below.

Task of quantity-number interference (CANUM, Gutiérrez-Martínez, Ramos-Ortega, & Vila, 2018). A group of numbers is shown on the screen. In the centre, there is a number in black that is presented one, two, three or four times. On each side of the black numbers, there is a number in grey (e.g., **14443**). One of these grey numbers corresponds to the number of times the central (black) number is presented. Participants are instructed to respond by clicking once the mouse button (right or left) on the side showing the number that corresponds to the number of times the central number is repeated (relevant dimension), and to ignore the numerical symbol (irrelevant dimension). In the example shown above, the right mouse button should be clicked because the central black number **4** is repeated three times, and the grey number **3** is presented to the right. As in the Stroop color-word test, the participant would face congruent (e.g., **2221,** "2 twos") or incongruous configurations (e.g., **12223**, "3 twos"). Therefore, facilitating (congruent cases) or interference (incongruous cases) effects can be generated. The task forces active maintenance, selection, and inhibition processes to effectively manage interference and preponderant response trends. To make the task applicable to 4–5-year-old children, 6 training and 48 (24 congruent, 24 incongruous) items were selected from the 120 items that constituted the task. One point was awarded for each correct trial (i.e., maximum score 48 points). Cronbach's alpha = .910.

Name-sound correspondence (PRIM, Gutiérrez-Martínez, & Vila, 2004). This is a task of image-sound correspondence used to measure executive-attentional-inhibitory ability. The task was limited to 60 trials (10 training trials and 50 test trials) but including the same categories: objects of eight categories animal, color, fruit, music, numbers, cloths, transport, and tools. An image is shown on the centre of the computer screen followed by a name sound. Participants are instructed to respond by clicking the left mouse button if the name and image correspond or the right mouse button if they differ. The task includes a small number of incorrect matches (5) to bias the answer. Therefore, although task processing is easy, it requires sustained attention. One point was awarded for each correct trial (i.e., maximum score 50 points). Cronbach's alpha = .910.

Working Memory Task (CATEG-WM, Gutiérrez-Martínez & Vila, 2004**).** This task forces the subject both to maintain attentional control of the double task (switching between processing and storage), and to "update" at the time of the response (control of possible interference between the

two categories). Images are presented in 2x2 matrices on the computer screen. Participants are instructed to complete two tasks simultaneously. The participant is asked to select and name the image that is different from the other three, the one that does not belong to the general category, *the intruder*. The name of the intruder has to be kept in mind because after some trials a big question mark will appear on the screen, and the participant will be required to name all the intruders in the same order as they were shown (i.e., recalling task). The task consists of three levels gradually increasing in the number of items to remember (intruders) (2 items in the first level, 3 items in the second level, and 4 items in the third level). To pass to the following level, the participant has to obtain at least one point on the previous level. Two points are awarded if the intruders are named in the correct order, one point if the child names all intruders, but in an incorrect order. The number of correct answers in each level were recorded, and multiplied by 2, 3 or 4 depending on the level. The maximum score for this task is 54 points. Cronbach's alpha = .841.

The stimuli were computerized and applied through LEEDUCA, an Internet-based platform that allows stimuli presentation and recording of both Reaction Time and Success-Errors. For the CANUM and PRIM tasks, correct-incorrect answers and the time taken to respond were recorded. For the CATEG-WM task, only correct answers were recorded.

Statistical analyses

First, evidence of internal structure (factor validity) of the CHEXI was examined. As previous studies have repeatedly found a two-factor structure for the CHEXI, we performed the Confirmatory Factor Analyses (*CFA*) to assess the adequacy of this two-factor model with WM as one of the factors and Inhibition as the other factor. The following criteria were used: 1) a χ^2/df value lower than 3 indicates a good fit; 2) a Comparative Fit Index (*CFI*; Bentler, 1990) greater than .95 constitute good fit between .90 and .95 show an acceptable fit, and values lower than .90 a poor fit; 3) a Root Mean Square Error of Approximation (*RMSEA*) value lower than .05 reflects a good fit, between .05 and .80 is a moderate fit, greater than .80 indicates a poor fit; 4) a Root Mean Square Residual (*RMSR*) value ranging from 0 to 1 represents a better model fit. An analysis of the measurement invariance with two factors (i.e., age and gender) was

also carried out. Internal consistency of the obtained factors was examined using Cronbach's alpha (α) and McDonald's omega (ω).

Second and temporal stability in parent ratings collected one year later was examined using bivariate correlations. Third, correlations were used to examine associations between the CHEXI subscales and achievement on the three EF tasks. Finally, A MANOVA was conducted to examine mean comparisons among the groups (girls vs. boys) and age (4 vs 5 years) on the CHEXI ratings.

In the final data matrix, 5.78% had missing data on items not answered or poorly filled in. No imputation of missing data was used." Only data from children who completed the three behavioural tests and whose parent completed the CHEXI ratings were included in the analyses.
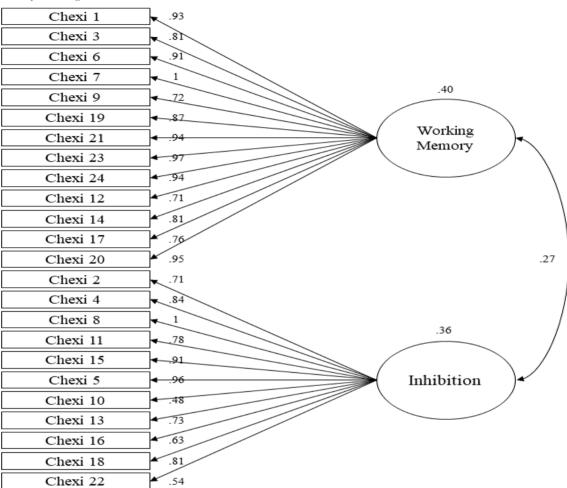
Results

Confirmatory factor analysis (CFA)

With regards to the first objective, the two-factor model originally established by Thorell and Nyberg (2008) was tested using *CFA* with maximum likelihood (*ML*) estimation. This model produced moderate-to-good fit indices: 1) a significant chi-square value, $\chi^2 (251) = 936.5, p < .00$; 2) a χ^2/df ratio of 373.1, which indicates a moderate fit; 3) an acceptable *CFI* of .913; 4) a *RMSR* of .044, which indicates a good fit to the data; 5) a *RMSEA* of .055, which also indicates a moderate model fit. Items and their factor loadings are presented in Figure 1 and they showed high correspondence with a two-factor solution. An analysis of the measurement invariance with two factors (i.e., age and gender) was also carried out. The fit of the model decreased when taking subgroups within the sample into account (see Table 1), but it remained acceptable according to the ratio relating to the degrees of freedom and the *RMSEA* (Cieciuch, & Davidov, 2015).

As shown in figure 1, except three items in the Inhibition subscale, items presented factor loadings over .70. The items with low loading were the following: 10-*Gets overly excited when something special is going to happen (e.g., going on a field trip, going to a party*; 16-Has difficulty refraining from smiling or laughing in situations where it is inappropriate, and 22-*Acts in a wilder way compared to other children in a group (e.g., at a birthday party or during a group activity)*.

Figure 1
Factor structure and factor loadings.

The two resulting subscales for the Spanish CHEXI showed good internal consistency for the 4 and 5 y-o samples, respectively. The WM subscale exhibited McDonald's ω-values of .89 and .89; and Cronbach's α-values of .89 and .92. The Inhibition subscale obtained ω-values of .82, and .84, and α-values of .78 and .84.

Table 1
Fit indices for invariance tests of the two-factor model (sample: n = 904).

Model	χ^2	df	p	CFI	RMSEA	Δ-CFI
Gender						
Configural invariance	109.2	416	.000	.913	.043	
Metric invariance	1148.0	461	.000	.912	.042	.001
Scalar invariance	1186.4	483	.000	.910	.041	.002
Age						
Configural invariance	1039.6	416	.000	.920	.041	
Metric invariance	1103.9	458	.000	.917	.040	.003
Scalar invariance	1178.6	483	.000	.911	.040	.006

Temporal stability

Pearson correlations were calculated to examine the temporal stability of the parent's ratings after one-year interval. The results showed that the temporal stability was high for the total score (r = .71 for 4-year-olds and r = .65 for 5-year-olds), as well as for both the WM subscale (r = .61 and r = .68, p < .001) and the Inhibition subscale (r = .61 and r = .63, p < .001).

Effects of age and gender

A MANOVA compared CHEXI total scores and subscales scores across gender (404 girls, 500 boys), and age groups (404 4-y-olds, 500 5-y-olds) as displayed in table 2. For age, participants differed in relation to CHEXI total scores and both subscales, with the 4-age group consistently scoring higher than the 5-age group. For WM subscale, $F(1,903) = 3.653$, $p = .056$, $\eta^2 p = .004$; the Inhibition sub-

scale, $F(1,903) = 11.288$, $p = .001$, $\eta^2 p = .012$; the Total scores, $F(1,903) = 8.342$, $p = .004$, $\eta^2 p = .009$.

Results revealed that girls and boys differed on the WM subscale, $F(1,903) = 5.700$, $p = .017$, $\eta^2 p = .006$; on the Inhibition subscale, $F(1,903) = 7.005$, $p = .008$, $\eta^2 p = .008$; on the Total scores, $F(1,903) = 7.771$, $p = .005$, $\eta^2 p = .009$. However, no significant interaction effects of age and gender were found.

Table 2
Mean and standard deviations for the CHEXI subscales Working Memory and Inhibition, total scores and mean comparisons by age and gender (MANOVA).

	Boys		Girls		F-value (η^2)	F-value (η^2)
	4 years, M (SD)	5-years, M (SD)	4-years, M (SD)	5 years, M (SD)	by Age	by Gender
WM	26.99 (8.3)	25.46 (7.6)	25.13 (7.8)	24.73 (8.5)	3.653* (.004)	5.700* (.006)
Inhibition	31.29 (7.1)	29.43 (7.3)	29.74 (7.6)	28.38 (7.4)	11.288** (.012)	7.005** (.008)
Total	58.28 (13.6)	54.88 (13.4)	54.86 (13.9)	53.10 (14.7)	8.342* (.009)	7.771* (.009)

*$p < .05$; **$p < .001$

Associations between the CHEXI and EF tasks

Table 3 presents correlations between the scores on the CHEXI scales and performance on EF tasks measuring either inhibitory control (i.e., PRIM and CANUM tasks) and WM (i.e., CANTEG-WM task). Among 4-year-olds, the CHEXI WM subscale was significantly related to performance on PRIM, CATEG-WM and CANUM-Time, but not

to performance on CANUM-CA. The Inhibition subscale was only significantly related to CANUM-Time. Among 5-year-olds, both the CHEXI WM subscale and the CHEXI Inhibition were significantly related to CANTUM-WM, but no other significant relations were found among 5-year-olds. It should also be noted that the highest correlation found was only $r = .15$.

Table 3
Mean Laboratory correct answers (CA) and Time (T) by age, correlations between the CHEXI factors Working Memory (WM) and Inhibition (INH), and results of ANOVA by age on the laboratory measures.

		4 years			5 years			F-value, by Age	
			CHEXI			CHEXI			
		Mean (SD)	WM	INH	Mean (SD)	WM	INH	F	p
CATEG-WM		10.6 (8.3)	-.11*	-.02	19.1 (11.3)	-.10*	-.09*	87.759	.000
CANUM	CA	31.4 (8.4)	-.04	-.06	33.6 (6.4)	-.05	-.02	23.147	.000
	Time	427.7 (159.3)	.15**	.12*	330.8 (105.3)	.08	.09	91.103	.000
PRIM	CA	13.3 (8.6)	-.15**	-.06	43.7 (5.6)	-.02	-.03	10.883	.001
	Time	-	-	-	123.4 (37.7)	.09	.07	-	-

*$p < .05$; **$p < .001$
Note. CA = Correct Answers, WM = Working Memory, INH = Inhibition

Discussion

The present study was aimed to investigate the psychometric properties of the Spanish version of the CHEXI, a rating scale designed to measure everyday EF in children. The findings indicated that the CHEXI shows adequate model fit indexes, high internal consistency of the obtained factors, as well as good temporal stability across a one-year time interval. With regard to effects of age and gender, the results showed that boys had larger EF deficits compared to girls and 4-year-olds had larger EF deficits compared to 5-year-olds, especially with regard to inhibition. Finally, the results showed that the CHEXI subscales were only weakly, although in some cases significantly, related to performance on EF tasks.

Factor structure of the Spanish CHEXI

The results confirmed the same two-factor model that emerged in the original version of the CHEXI presented by Thorell & Nyberg (2008), as well as in subsequent adaptations in different languages (e.g., Camerota et al., 2018; Catale et al., 2013; Tonietti et al., 2017). Internal consistency of the two factors was also satisfactory, as was the temporal stability. Thus, parents' ratings reflect that the two major EF components, WM and Inhibition, constitute distinct aspects of children's cognitive functioning at an early age. An examination of the Inhibition subscale content revealed that the three items with poor loadings on this factor explore overacting in social situations. This finding is important since it might indicate that preschoolers show signs of different EF components as a function of context. Based on these find-

ings, we consider that the Spanish version of the CHEXI provide reliable measures of everyday EF in preschoolers.

Effects of age and gender

With regard to effects of age, 5-year-old children were rated as having better executive control than 4-year-olds and were better performers on the EF tasks. These findings are consistent with previous research that point to a rapid development in EF during the preschool age (e.g., Garon, Bryson, & Smith, 2008). Interestingly, the age difference was especially evident for the CHEXI Inhibition subscale, which could be taken to indicate that EF components evolve at different speed. For example, Simpson and Riggs (2005) found that 3.5 and 5-year-old children differed in their speed and accuracy in the 'night and day task' measuring inhibition, and that a reduction in the memory load did not result in improved performance. These and other similar results provide evidence that children's inhibitory control show quick improvements at the preschool age, while WM might develop more gradually (for a review, see Best & Miller, 2010). As suggested by Diamond et al. (2002), it may sometimes happen that children know the correct answer, although they are unable to inhibit the wrong answer.

When examining gender differences, boys were rated as having more problems with executive control than girls with regard to both WM and Inhibition. Our results are comparable to prior findings of a female EF advantage in the CHEXI validation study for US preschoolers (Camerota et al., 2018) as well as in a CHEXI cross-cultural study of children aged 6-11 years (Thorell et al., 2013). Using the Behavior Rating Inventory of Executive Function (BRIEF)**,** Sherman and Brooks (2010), showed that parents rated boys aged 2 to 5 years as having somewhat poorer inhibitory control than girls. Interestingly, this study showed that, contrary to the gender difference found in parents' ratings, no gender differences were found for EF laboratory tasks. Yamamoto & Imai-Matsumura (2019) found the same discrepant results between ratings and direct measures. Altogether, these findings suggest that gender differences in ratings might be better explained by the boy's more elevated levels of externalizing behaviors or the influence of cultural patterns rather than an actual lower EF ability.

A related concern is that parents might not be accurate raters because they lack the teachers' experience or interact with a reduced number of children of the same age (Korsch & Petermann, 2014). However, Thorell and Nyberg, (2008) and Thorell et al., (2010) obtained the same two-factor structure from parent and teacher ratings on the CHEXI, indicating that both parents and teachers can differentiate between WM and Inhibition in children.

Associations between EF ratings and EF tasks

The weak associations found between CHEXI ratings and the EF tests replicate results from several previous stud-

ies (Camerota et al., 2018; Catale et al., 2013; Tonietti et al., 2017). As suggested by Toplak et al. (2013) this could be a result of the fact that EF tasks and EF ratings capture at least partly different constructs. Nevertheless, the small but significant association in both age groups between children's performance on the WM task (CATEG-WM) and the CHEXI ratings indicate that both measures capture some common traits. Rather than concluding that ratings are better than tests or vice versa, we believe that different types of EF measures should be seen as complimentary to one another. EF tasks are appropriate to assess optimal performance on specific components of EF during a limited time, while ratings capture average EF skills over an extended period of time, and provide a better view of daily life problems related to EF deficits (Anderson, 2002; Toplak et al., 2013). We therefore believe that the CHEXI should be regarded as a valuable measure of preschoolers' EF functioning in natural settings.

Strengths, limitations and conclusions

This study was characterized by three key strengths. First, it included a large sample from diverse sociodemographic backgrounds. Given the high degree of variability in self-regulation among preschool-age children, this could be taken to suggest that our results can be generalized to a broader population. Second, we found temporal stability over a longer period than previous studies, which increase our confidence in that the scores obtained from the CHEXI provides a reliable measure of EF in young children. Third, the Spanish CHEXI was shown to have high internal consistency and the same two-factor structure found in previous studies could be replicated.

The limitations of this study arise first, from the fact that it was not included a clinical sample. Further work needs to examine to what extent the Spanish CHEXI can differentiate between normally developing children and children with disorders known to be associated with EF deficits (e.g., those with ADHD). Second, it was no registered which parent (mother or father) answered the questionnaire, neither if the same parent responded the second time. Evidence advises that there might be differences among raters. Third, unpublished EF tasks were used. However, these tasks were adaptations of tasks that have successfully been used in the assessment of EF deficits in school-aged children (Gutiérrez-Martínez et al., 2018).

In sum, the present study supports and extends previous research on the measurement of EF in preschool-age children by showing that the scores obtained from the Spanish CHEXI are reliable and provides a suitable instrument to assess EF in young children. However, due to the low association between the CHEXI and EF test, ratings such as the CHEXI should preferably be used as a screening measure or in combination with EF tests in order to obtain a more detailed picture of a child's EF ability.

References

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*, 71-82. DOI: 10.1076/chin.8.2.71.8724

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin, 107*, 238-246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Best, J. & Miller, P. (2010). A Developmental Perspective on Executive Function. *Child Dev., 81,* 1641-1660. doi: 10.1111/j.1467-8624.2010.01499.x

Best, J.R., Miller, P.H., & Jones, L.L. (2009). Executive Functions after Age 5: Changes and Correlates. *Developmental Review, 29*, 180-200. https://doi.org/10.1016/j.dr.2009.05.002

Best, J., Miller, P., & Naglieri, J. (2011). Relations between Executive Function and Academic Achievement from Ages 5 to 17 in a Large, Representative National Sample. *Learning Individual Differences, 21*: 327-336. DOI: 10.1016/j.lindif.2011.01.007

Blair, C. & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology. 20*: 899-911. doi:10.1017/S0954579408000436

Camerota, M., Willoughby, M., Kuhn, L., & Blair, C. (2018). The Childhood Executive Functioning Inventory (CHEXI): Factor structure, measurement invariance, and correlates in US pre-schoolers, *Child Neuropsychology, 24,* 322-337. https://doi.org/10.1080/09297049.2016.1247795

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616. https://doi.org/10.1207/s15326942dn2802_3

Catale, C., Lejeune, C., Merbah, S., & Meulemans, Th. (2013). French Adaptation of the Childhood Executive Functioning Inventory (CHEXI). Confirmatory Factor Analysis in a Sample of Young French-Speaking Belgian Children. *European Journal of Psychological Assessment, 29,* 149–155. https://doi.org/10.1027/1015-5759/a000141

Catale, C., Meulemans, T., & Thorell, L.B. (2015). The Childhood Executive Function Inventory: Confirmatory factor analyses and cross-cultural clinical validity in a sample of 8- to 11-year-old children. *Journal of Attention Disorders, 19*, 489-495. https://doi.org/10.1177/1087054712470971

Cieciuch, J., & Davidov, E. (2015). Establishing measurement invariance across online and offline samples. A tutorial with the software packages Amos and Mplus. *Studia Psychologica: Theoria et praxis, 2*, 83-99.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*, 135-68. doi: 10.1146/annurev-psych-113011-143750

Diamond, A. (2020). Executive functions. In J.L. Michaud, C. Bulteau, D. Cohen, & A. Gallagher (Eds.), *Handbook of Clinical Neurology, 173*, 225-240. Amsterdam: Elsevier. ISBN:978-0444641502

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*, 352–362. doi: 10.1037/0012-1649.38.3.352

Friedman, N.P., Haberstick, B.C., Willcutt, E.G., Miyake, A., Young, S.E. et al. (2007). Greater Attention Problems During Childhood Predict Poorer Executive Functioning in Late Adolescence. *Psychological Science, 18,* 893-900. doi: 10.1111/j.1467-9280.2007.01997.x

García-Madruga, J.A., Gómez-Veiga, I., & Vila, J.O. (2016). WM executive processes and the improvement of thinking abilities: The intervention in reading comprehension. *Frontiers in Psychology, 7,* 58. https://doi.org/10.3389/fpsyg.2016.00058

García-Villamisar, D., Jodra-Chuan, M., Saez, G., & Thorell, L.B. (2020). *Spanish version of the Adult Executive Functioning Inventory (ADEXI): Psychometric Properties in Adults with Autism Spectrum Disorders. Journal of Psychopathology and Clinical Psychology, 25,* 121-129. doi: 10.5944/rppc.26183

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Review, 134,* 31-60. https://doi.org/10.1037/0033-2909.134.1.31

Gutiérrez-Martínez, F., & Vila, J. O. (2004). Prueba de correspondencia nominal (léxico−perceptiva): PRIM y Prueba de Amplitud de Categorías: CATEG_A. Unpublished manuscript.

Gutiérrez-Martínez, F., Ramos-Ortega, M., & Vila, J. O. (2018). Executive efficacy on Stroop type interference tasks. A validation study of a numerical and manual version (CANUM). *Anales de Psicología, 34,* 184-196. https://doi.org/10.6018/analesps.34.1.263431

Holst, Y. & Thorell, L.B. (2018). Adult Executive Functioning Inventory (ADEXI): Validity, Reliability, and Relations to ADHD. *International Journal of Methods in Psychiatric Research, 27:*e1567. https://doi.org/10.1002/mpr.1567

Isquith, P., Gioia, G., & Espy, K. (2004). Executive function in preschool children: examination through everyday behaviour. *Developmental Neuropsychology, 26*, 403-422. https://doi.org/10.1207/s15326942dn2601_3

Korsch, F. & Petermann, F. (2014). Agreement between parents and teachers on preschool children's behavior in a clinical sample with externalizing behavioral problems. *Child Psychiatry Hum. Dev., 45,* 617–27. doi 10.1007/s10578-013-0430-6

Miyake, A. & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Curr Dir Psychol Sci., 21,* 8–14. doi: 10.1177/0963721411429458

Moffitt, T.E., Arseneault, l, Belsky, D., Dickson, N., Hancox, R. et al. (2011). A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety. *Proceedings of the National Academy of Sciences 108*, 2693-8. doi: 10.1073/pnas.1010076108

Romine, C.B. & Reynolds, C.R. (2005). A Model of the Development of Frontal Lobe Functioning: Findings from a meta-analysis. *Applied Neuropsychology, 12,* 190-201. doi: https://doi.org/10.1207/s15324826an1204_2

Shaul, S., & Schwartz, M. (2014). The role of the executive functions in school readiness among preschool-age children. *Reading & Writing, 27,* 749.

Sherman, E. & Brooks, B.L. (2010). Behavior rating inventory of executive function – preschool version (BRIEF-P): Test review and clinical guidelines for use. *Child Neuropsychology, 16,* 503-519. https://doi.org/10.1080/09297041003679344

Simpson, A. & Riggs, K.J. (2005). Inhibitory and working memory demands of the day–night task in children. *British Journal of Developmental Psychology, 23*, 471–486. https://doi.org/10.1348/026151005X28712

Sjöwall, D., Bohlin, G., Rydell, A.M., & Thorell, L. (2017). Neuropsychological deficits in preschool as predictors of ADHD symptoms and academic achievement in late adolescence. *Child Neuropsychology, 23*, 111-128. https://doi.org/10.1080/09297049.2015.1063595

Thorell, L.B. & Catale, C. (2014). The assessment of executive functioning using the Childhood Executive Function Inventory (CHEXI). S. Goldstein and J. Naglieri (ed). *Handbook of executive functioning.* New York: Springer. Pp. 359-366. DOI: 10.1007/978-1-4614-8106-5

Thorell, L.B. & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A New Rating Instrument for Parents and Teachers. *Developmental Neuropsychology, 33,* 536-552. https://doi.org/10.1080/87565640802101516

Thorell, L.B., Eninger, L., Brocki, K.C. & Bohlin, G. (2010). Childhood Executive Function Inventory (CHEXI): A promising measure for identifying young children with ADHD? *Journal of Clinical and Experimental Neuropsychology, 32,* 38-43. https://doi.org/10.1080/13803390902806527

Thorell, L.B., Veleiro, A., Siu, A.F., & Mohammadi, H. (2013). Examining the relation between ratings of executive functioning and academic achievement: Findings from a cross-cultural study. *Child Neuropsychology, 19*, 630-638. https://doi.org/10.1080/09297049.2012.727792

Tonietti, B., Martins, N., de Almeida, A., & Gotuzo, A. (2017). Childhood Executive Functioning Inventory: Adaptação e Propriedades Psicomé-

tricas da Versão Brasileira. *Psico-USF, Bragança Paulista, 22*, 63-74. http://dx.doi.org/10.1590/1413-82712017220106

Toplak, M., West, M., & Stanovich, R. (2013). Practitioner review do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, 54*, 131–143. https://doi.org/10.1111/jcpp.12001

Willcutt, E.G., Doyle, A.E., Nigg, J.T., Faraone, S.V. & Pennington, B.F. (2005). Validity of the Executive Function Theory of Attention-

Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biol Psychiatry, 57*, 1336–1346. doi:10.1016/j.biopsych.2005.02.006

Yamamoto, N. & Imai-Matsumura, K. (2019) Gender differences in executive function and behavioural self-regulation in 5 years old kindergarteners from East Japan. *Early Child Development and Care, 189*, 56-67. https://doi.org/10.1080/03004430.2017.1299148